A RAPIDLY-CONVERGING LOWER BOUND FOR THE JOINT SPECTRAL RADIUS VIA MULTIPLICATIVE ERGODIC THEORY

IAN D. MORRIS

ABSTRACT. We use ergodic theory to prove a quantitative version of a theorem of M. A. Berger and Y. Wang, which relates the joint spectral radius of a set of matrices to the spectral radii of finite products of those matrices. The proof rests on a theorem asserting the existence of a continuous invariant splitting for certain matrix cocycles defined over a minimal homeomorphism and having the property that all forward products are uniformly bounded. MSC primary 15A18, 37H15, 65F15, secondary 37M25.

1. INTRODUCTION

Let A be a bounded set of $d \times d$ complex matrices. The joint spectral radius of A, introduced by G.-C. Rota and G. Strang in [41], is defined to be the quantity

$$(1) \qquad \varrho(\mathsf{A}) := \lim_{n\to\infty} \sup \left\{ \|A_n \cdots A_1\|^{1/n} \colon A_i \in \mathsf{A} \right\},$$

where $\| \cdot \|$ denotes any norm on $\mathbb{C}^d$. This is easily seen to yield a finite value which is well-defined with respect to the choice of norm. The joint spectral radius arises naturally in a range of topics including control and stability [1, 24, 31], coding theory [36], wavelet regularity [15, 16, 35], numerical solutions to ordinary differential equations [23], and combinatorics [17]. The problem of computing the joint spectral radius of a finite set of matrices has therefore attracted substantial research interest [3, 21, 22, 31, 33, 34, 39, 46, 48]. In this article we shall prove a new estimate relevant to the computation of the joint spectral radius.

Let $\mathrm{Mat}_d(\mathbb{C})$ denote the set of all $d \times d$ complex matrices. The following theorem was proved by M. A. Berger and Y. Wang [2], having originally been conjectured by I. Daubechies and J. C. Lagarias [15]:

Theorem 1.1 (Berger-Wang formula). *Let $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ be bounded. Then*

$$(2) \qquad \varrho(\mathsf{A}) = \limsup_{n\to\infty} \sup \left\{ \rho(A_n \cdots A_1)^{1/n} \colon A_i \in \mathsf{A} \right\},$$

where $\rho(A)$ denotes the ordinary spectral radius of a matrix A.

Some alternative proofs are given in [5, 18, 44]. In this article we shall study the rate of convergence in the expression (2). This has potential implications for some approaches to the computation of the joint spectral radius such as the algorithm given by G. Gripenberg [21].

Let $\| \cdot \|$ be any norm on $\mathbb{C}^d$. For each $n \in \mathbb{N}$ define

$$\varrho_n^+(\mathsf{A}, \| \cdot \|) = \sup \left\{ \|A_n \cdots A_1\|^{1/n} \colon A_i \in \mathsf{A} \right\},$$

$$\varrho_n^-(\mathsf{A}) = \sup\left\{\rho(A_n \cdots A_1)^{1/n} : A_i \in \mathsf{A}\right\}.$$

For fixed A it is clear that $\varrho_{n+m}^+(\mathsf{A}, \|\cdot\|) \leq \varrho_n^+(\mathsf{A}, \|\cdot\|)\varrho_m^+(\mathsf{A}, \|\cdot\|)$ for all $n, m \in \mathbb{N}$, which implies that the limit in (1) may be replaced by an infimum. Conversely, since $\rho(A^m)^{1/m} = \rho(A)$ for all $m \in \mathbb{N}$ and any matrix A, one may easily show that $\varrho_{nm}^-(\mathsf{A}) \geq \varrho_n^-(\mathsf{A})$ for every $n, m \in \mathbb{N}$ and hence the limit superior in (2) is also a supremum. In general this limit superior can fail to be a limit, a simple example being

$$\mathsf{A} = \left\{\begin{pmatrix} 0 & 2 \\ \frac{1}{2} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right\}.$$

In this article we shall present a proof of the following theorem, which extends Theorem 1.1 in the case where A is finite:

Theorem 1.2. *Let A be a finite set of $d \times d$ complex matrices. Then for every integer $r \in \mathbb{N}$,*

$$\left|\varrho(\mathsf{A}) - \max_{1 \leq k \leq n} \varrho_k^-(\mathsf{A})\right| = O\left(\frac{1}{n^r}\right).$$

Theorem 1.2 implies in particular that if we wish to compute $\varrho(\mathsf{A})$ to within accuracy ε by means of brute-force estimation of the values $\varrho_n^-(\mathsf{A})$, then the number of matrix products which must be evaluated increases at a slower-than-stretched-exponential rate as a function of $1/\varepsilon$. However, it should be noted that the arguments used in this paper do not seem to be well-suited to the production of an *effective* estimate for the quantity $\varrho(\mathsf{A})$.

Two estimates related to Theorem 1.2 have been established previously. By a theorem of J. Bochi [5], there exist for each $d \in \mathbb{N}$ a constant $C_d > 0$ and an integer $m \in \mathbb{N}$ such that $\varrho(\mathsf{A}) \leq C_d \max_{1 \leq k \leq m} \varrho_k^-(\mathsf{A})$ for every bounded set $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$. An easy consequence is the estimate

$$\left|\varrho(\mathsf{A}) - \max_{1 \leq k \leq mn} \varrho_k^-(\mathsf{A})\right| \leq \left(1 - C_d^{-1/n}\right)\varrho(\mathsf{A}) = O\left(\frac{1}{n}\right).$$

In the other direction, F. Wirth [48] gives the general bound

$$\left|\varrho(\mathsf{A}) - \varrho_n^+(\mathsf{A}, \|\cdot\|)\right| = O\left(\frac{\log n}{n}\right)$$

for any norm $\|\cdot\|$ on $\mathbb{C}^d$ and bounded set $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$. This estimate improves to $O(1/n)$ if it is assumed that there does not exist a linear space V such that $\{0\} \subset V \subset \mathbb{C}^d$ and $AV \subseteq V$ for every $A \in \mathsf{A}$. Unlike Bochi's estimate, the constant in Wirth's estimate may vary between sets of matrices A. The example

$$\mathsf{A} = \left\{\begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right\}$$

shows that Wirth's estimate cannot be improved directly: taking $\|\cdot\|$ to be the Euclidean norm we obtain $\varrho_n^+(\mathsf{A}, \|\cdot\|) = 2^{1+1/2n}$ for each $n \in \mathbb{N}$, whereas $\varrho_1^-(\mathsf{A}) = 2$ and hence $\varrho(\mathsf{A}) = 2$.

The proof of Theorem 1.2 has some points of resemblance to the proof of Theorem 1.1 given by L. Elsner [18], which we now elaborate upon. Elsner's proof runs essentially as follows. If $\varrho(\mathsf{A}) = 0$ then the result is trivially true. Otherwise, by normalising we may take $\varrho(\mathsf{A}) = 1$. We then reduce to the case where a uniform bound exists for products of elements of A, and hence there exists a compact subset of $\mathrm{Mat}_d(\mathbb{C})$ which contains $\{A_n \ldots A_1 : A_i \in \mathsf{A}\}$ for every n. By using the pigeonhole

principle on open ε-balls in $\mathrm{Mat}_d(\mathbb{C})$ and in $\mathbb{C}^d$, we can then guarantee the existence of a finite sequence $A_1, \ldots, A_n$ and a vector v belonging to the unit sphere of $\mathbb{C}^d$ such that $A_n \cdots A_1 v$ is close to v and therefore the spectral radius of $A_n \cdots A_1$ is close to 1.

In our proof of Theorem 1.2 we make this strategy quantitative, replacing the pigeonhole principle with a more delicate recurrence argument. In order to achieve this we first prove a theorem describing the dynamical structure of matrix sequences (A_i) with the property that $\|A_n \cdots A_1\|$ is large for all n, and additionally we achieve some understanding of the structure of the orbits in $\mathbb{C}^d$ which are induced by the action of such sequences. The bulk of this paper, therefore, is concerned with proving a theorem on the dynamical structure of these 'extremal' sequences. We describe these ideas in detail in the following section.

2. Linear cocycles

At this point it is convenient to establish some notation and definitions. In the remainder of this article the symbol $\|\cdot\|$ shall be used to denote the Euclidean norm on $\mathbb{C}^d$, whereas the symbol $\|\cdot\|$ shall be used to denote an *extremal norm* on $\mathbb{C}^d$, which will be defined shortly. In either case we shall also use the symbols $\|\cdot\|$ and $\|\cdot\|$ to denote the corresponding operator norms induced on $\mathrm{Mat}_d(\mathbb{C})$. Throughout this article we adhere to the convention $\log 0 := -\infty$.

Let $T \colon X \to X$ be a continuous transformation of a compact metric space. A *cocycle* over T with values in the complex matrices is a function $\mathcal{A} \colon X \times \mathbb{N} \to \mathrm{Mat}_d(\mathbb{C})$ such that for each $x \in X$ and $n, m \in \mathbb{N}$

$$\mathcal{A}(x, n + m) = \mathcal{A}(T^n x, m)\mathcal{A}(x, m).$$

We say that the cocycle $\mathcal{A}$ is continuous if $\mathcal{A}(\cdot, n)$ is a continuous function from X to $\mathrm{Mat}_d(\mathbb{C})$ for each $n \in \mathbb{N}$. Abusing notation somewhat, we shall sometimes denote $\mathcal{A}(x, 1)$ simply by $\mathcal{A}(x)$. Since for each x, n

$$\mathcal{A}(x, n) = \mathcal{A}(T^{n-1}x) \cdots \mathcal{A}(Tx)\mathcal{A}(x)$$

the cocycle $\mathcal{A} \colon X \times \mathbb{N} \to \mathrm{Mat}_d(\mathbb{C})$ is completely determined by the function $\mathcal{A} \colon X \to \mathrm{Mat}_d(\mathbb{C})$. Whilst it will always be the case in this article that the map T is a homeomorpism, we do not assume that the values of the function $\mathcal{A}$ are invertible matrices, and so we cannot in general extend $\mathcal{A}$ to an invertible cocycle defined on $X \times \mathbb{Z}$.

For $0 \leq p \leq d$ we let $\mathrm{Gr}(p, d)$ denote the set of all p-dimensional subspaces of $\mathbb{C}^d$. This set may be identified with the set of all orthogonal projections from $\mathbb{C}^d$ onto a p-dimensional subspace. We equip $\mathrm{Gr}(p, d)$ with the standard metric given by

$$d_{\mathrm{Gr}}(V, W) := \|P_V^\perp - P_W^\perp\|$$

where $P_Z^\perp$ denotes the linear map given by orthogonal projection onto Z. This metric makes $\mathrm{Gr}(p, d)$ a compact metric space. We shall say that a function $\mathcal{V} \colon X \to \mathrm{Gr}(p, d)$ is *forward-invariant* under a cocycle $\mathcal{A}$ if $\mathcal{A}(x, n)\mathcal{V}(x) \subseteq \mathcal{V}(T^n x)$ for all $x \in X$ and $n \in \mathbb{N}$.

We begin by establishing the following general theorem which will later be applied to study matrix cocycles associated to a compact set $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$.

Theorem 2.1. *Let $T \colon X \to X$ be a minimal homeomorphism of a compact metric space, and let $\mathcal{A} \colon X \times \mathbb{N} \to \mathrm{Mat}_d(\mathbb{C})$ be a continuous linear cocycle. Suppose*

that there exists $M > 0$ such that $\|\mathcal{A}(x, n)\| \leq M$ for all $x \in X$ and all $n \in \mathbb{N}$. Then there exist an integer $0 \leq p \leq d$ and continuous forward-invariant functions $\mathcal{V} \colon X \to \mathrm{Gr}(p, d)$, $\mathcal{W} \colon X \to \mathrm{Gr}(d-p, d)$ such that $\mathcal{V}(x) \oplus \mathcal{W}(x) = \mathbb{C}^d$ for all $x \in X$. Moreover there exist constants $C, \delta > 0$ and $\xi \in (0, 1)$ such that for all $x \in X$ and $n \in \mathbb{N}$, $\|\mathcal{A}(x, n)v\| \geq \delta$ for every $v \in \mathcal{V}(x)$ and $\|\mathcal{A}(x, n)w\| \leq C\xi^n \|w\|$ for every $w \in \mathcal{W}(x)$.

The moduli of continuity of $\mathcal{V}$ and $\mathcal{W}$ admit the following description. If $n \in \mathbb{N}$ is given, suppose that $x, y \in X$ satisfy

$$\max \left\{ \|\mathcal{A}(x, 2n) - \mathcal{A}(y, 2n)\|, \|\mathcal{A}(x, n) - \mathcal{A}(y, n)\| \right\} \leq \delta\xi^n.$$

Then $d_{\mathrm{Gr}}(\mathcal{V}(x), \mathcal{V}(y)) \leq \tilde{C}\xi^n$ for some constant $\tilde{C} > 0$. Similarly, if $x, y \in X$ satisfy

$$\|\mathcal{A}(x, n) - \mathcal{A}(y, n)\| \leq \xi^n$$

then $d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y)) \leq \tilde{C}\xi^n$.

For each $x \in X$ let $P(x)$ denote the projection with image $\mathcal{V}(x)$ and kernel $\mathcal{W}(x)$. Then $P(x)$ depends continuously on x, and in particular there exists $K > 0$ such that

$$\|P(x) - P(y)\| \leq K \left[d_{\mathrm{Gr}}(\mathcal{V}(x), \mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y)) \right]$$

for all $x, y \in X$.

While Theorem 2.1 has a number of features in common with the classical multiplicative ergodic theorem of V. I. Oseledec (see e.g. [32]) our proof is direct and does not make use of any prior multiplicative ergodic theorems. Indeed, since in general we wish to work with non-invertible matrices, the standard statement of Oseledec's theorem does not give the existence even of a measurable splitting of the type given above, giving only an invariant flag (though see [19]). The proof of Theorem 2.1 does however incorporate ideas used in the proofs of Oseledec's theorem given by M. S. Raghunathan [40] and D. Ruelle [42].

Note that if $p = 0$ then the conclusions of the theorem are somewhat vacuous, and in applications further analysis is needed to show that this situation does not arise.

In order to apply this theorem in the desired context we require some further definitions. We shall say that $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ is *product bounded* if there exists $M > 0$ such that for every $n \in \mathbb{N}$ we have $\|A_n \cdots A_1\| \leq M$ for every finite sequence $(A_n, \ldots, A_1) \in \mathsf{A}^n$. Note if such a uniform bound holds for A with respect to some norm on $\mathbb{C}^d$ then it holds with respect to all such norms, subject to variation in the constant M. We shall say that a norm $\|\!\cdot\!\|$ on $\mathbb{C}^d$ is an *extremal norm* for A if $\|\!A\!\| \leq \varrho(\mathsf{A})$ for all $A \in \mathsf{A}$. If $\varrho(\mathsf{A}) > 0$ then an extremal norm exists for A if and only if $\varrho(\mathsf{A})^{-1}\mathsf{A}$ is product bounded [31, 41].

Given a compact set $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$, let us define a metric on $\mathsf{A}^{\mathbb{Z}}$ by

$$d\left[(A_i)_{i \in \mathbb{Z}}, (B_i)_{i \in \mathbb{Z}}\right] := \sum_{i \in \mathbb{Z}} \frac{\|A_i - B_i\|}{2^{|i|}}.$$

If A is compact then $(\mathsf{A}^{\mathbb{Z}}, d)$ is compact. We define the *shift map* $T \colon \mathsf{A}^{\mathbb{Z}} \to \mathsf{A}^{\mathbb{Z}}$ by $T[(A_i)_{i \in \mathbb{Z}}] = (A_{i+1})_{i \in \mathbb{Z}}$. The shift map is a Lipschitz homeomorphism of $\mathsf{A}^{\mathbb{Z}}$. Let $\mathcal{A} \colon \mathsf{A}^{\mathbb{Z}} \to \mathrm{Mat}_d(\mathbb{C})$ be given by $\mathcal{A}[(A_i)_{i \in \mathbb{Z}}] = A_1$, and let $\mathcal{A}(x, n) = \mathcal{A}(T^{n-1}x) \cdots \mathcal{A}(x)$ for all $(x, n) \in \mathsf{A}^{\mathbb{Z}} \times \mathbb{N}$ so that $\mathcal{A} \colon \mathsf{A}^{\mathbb{Z}} \times \mathbb{N} \to \mathrm{Mat}_d(\mathbb{C})$ is a

continuous cocycle. For each $n \in \mathbb{N}$ we have

$$\varrho_n^+(\mathsf{A}, \|\cdot\|) = \sup\left\{\|\mathcal{A}(x,n)\|^{1/n} : x \in \mathsf{A}\right\}$$

and

$$\varrho_n^-(\mathsf{A}) = \sup\left\{\rho(\mathcal{A}(x,n))^{1/n} : x \in \mathsf{A}\right\}.$$

As a consequence we deduce

$$\log\varrho(\mathsf{A}) = \lim_{n\to\infty} \sup_{x\in\mathsf{A}^{\mathbb{Z}}} \frac{1}{n} \log\|\mathcal{A}(x,n)\|,$$

a formulation which is particularly amenable to study using ergodic theory via Theorem 3.4 below.

Combining Theorem 2.1 with some supplementary results given in section 3 below, we obtain the following:

Theorem 2.2. *Let $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ be a compact set such that $\varrho(\mathsf{A}) = 1$, and suppose that A is product bounded. Let $\|\cdot\|$ be any extremal norm for A and define*

$$Y := \left\{x \in \mathsf{A}^{\mathbb{Z}} : \|\mathcal{A}(x,n)\| = 1 \ \forall \ n \in \mathbb{N}\right\}.$$

Then the set Y is a compact, nonempty subset of $\mathsf{A}^{\mathbb{Z}}$ such that $TY \subseteq Y$.

Let $Z \subseteq Y$ be any invariant subset such that $T\colon Z \to Z$ is minimal. Then there exists an integer $1 \le p \le d$ such that the following properties hold. There exist Hölder continuous invariant functions $\mathcal{V}\colon Z \to \mathrm{Gr}(p,d)$, $\mathcal{W}\colon Z \to \mathrm{Gr}(d-p,d)$ such that $\mathcal{V}(x) \oplus \mathcal{W}(x) = \mathbb{C}^d$ for each $x \in Z$. There exist constants $C > 0$, $\xi \in (0,1)$ such that for all $x \in Z$ and $n \in \mathbb{N}$, $\|\mathcal{A}(x,n)v\| = \|v\|$ for all $v \in \mathcal{V}(x)$ and $\|\mathcal{A}(x,n)w\| \le C\xi^n\|w\|$ for all $w \in \mathcal{W}(x)$. If for each $x \in Z$ we let $P(x)$ denote the projection with image $\mathcal{V}(x)$ and kernel $\mathcal{W}(x)$ then $P\colon Z \to \mathrm{Mat}_d(\mathbb{C})$ is Hölder continuous.

To obtain Theorem 1.2 we combine this result with an estimate due to X. Bressaud and A. Quas on the approximation via periodic orbits of closed invariant subsets of shift transformations over finite alphabets (cf. [11]).

The remainder of this article is structured as follows. In section 3 we establish some results in subadditive ergodic theory which are needed in the proofs of Theorems 2.1 and 2.2. In sections 4 and 5 we prove these two theorems, and in section 6 we give the proof of Theorem 1.2. Finally, in section 7 we describe the obstructions to improving the error term in Theorem 1.2 and to extending that theorem to the case of infinite compact sets A.

3. Subadditive ergodic optimisation

The recently-developed topic of *ergodic optimisation* is concerned with the following problem. Given a continuous dynamical system $T\colon X \to X$ defined on a compact metric space, and some continuous (or only upper semi-continuous) function $f\colon X \to \mathbb{R}$, one studies the greatest possible linear growth rate of the sequence $\sum_{j=0}^{n-1} f(T^j x)$ as x varies over X, which is equal to the supremum of all possible values of the integral of f with respect to a T-invariant probability measure on X. Problems which are considered include the identification and approximation of those invariant measures which attain this supremum. Some recent research articles in this area include [7, 8, 10, 11, 13, 14, 28, 49]. In this section we generalise some (mostly standard) results from ergodic optimisation to the context of subadditive

ergodic theory, with the aim of applying these results to the proof of Theorem 2.1. For parallels of these results in the additive case we direct the reader to [28].

Throughout this section we assume that X is a compact metric space and $T\colon X \to X$ a continuous transformation. We let $\mathcal{M}$ denote the set of all Borel probability measures on X and let $\mathcal{M}_T$ denote the subset consisting of all T-invariant Borel probability measures. We equip $\mathcal{M}$ and $\mathcal{M}_T$ with the weak-* topology, under which both sets are compact and metrisable [47].

We in fact only require the results established below in the case where $f\colon X \to \mathbb{R} \cup \{-\infty\}$ is continuous, but the case in which f is only taken to be upper semi-continuous is included also since this does not require any modification to the proofs. The following simple result is important enough to be worth stating explicitly:

Lemma 3.1. *Let $f\colon X \to \mathbb{R} \cup \{-\infty\}$ be upper semi-continuous. Then the map from $\mathcal{M}$ to $\mathbb{R} \cup \{-\infty\}$ defined by $\mu \mapsto \int f\, d\mu$ is upper semi-continuous.*

Proof. Recall that a function from a metrisable space to $\mathbb{R} \cup \{-\infty\}$ is upper semi-continuous if and only if it is equal to the pointwise limit of a decreasing sequence of continuous functions taking values in $\mathbb{R}$ (see e.g. [6, ch. IX]). Let $(f_i)_{i=1}^{\infty}$ be such a sequence converging pointwise to f. For each i the map $\mu \mapsto \int f_i\, d\mu$ is clearly real-valued and is by definition weak-* continuous, and for each μ the sequence $(\int f_i\, d\mu)_{i=1}^{\infty}$ decreases to $\int f\, d\mu$ by the Monotone Convergence Theorem. $\square$

Recall that a sequence $(a_n)_{n=1}^{\infty}$ such that $a_n \in \mathbb{R} \cup \{-\infty\}$ for each n is said to be *subadditive* if $a_{n+m} \le a_n + a_m$ for all $n, m \in \mathbb{N}$. If this is the case then

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\ge 1} \frac{a_n}{n} \in \mathbb{R} \cup \{-\infty\}.$$

Definition 3.2. *We say that a sequence $(f_n)_{n=1}^{\infty}$ of functions from X to $\mathbb{R}\cup\{-\infty\}$ is* subadditive *if $f_{n+m}(x) \le f_n(T^m x) + f_m(x)$ for all $n, m \in \mathbb{N}$ and all $x \in X$.*

If $\mu \in \mathcal{M}_T$ and $(f_n)_{n=1}^{\infty}$ is a subadditive sequence of upper semi-continuous functions then the sequence $(\int f_n\, d\mu)_{n=1}^{\infty}$ is easily seen to be subadditive. If in addition $\mu \in \mathcal{M}_T$ is ergodic, then the Subadditive Ergodic Theorem asserts that for μ-a.e. $x \in X$

$$\lim_{n\to\infty} \frac{1}{n} f_n(x) = \lim_{n\to\infty} \frac{1}{n} \int f_n\, d\mu = \inf_{n\ge 1} \frac{1}{n} \int f_n\, d\mu,$$

see e.g. [32]. This motivates the following definition.

Definition 3.3. *Let (f_n) be a subadditive sequence of upper semi-continuous functions from X to $\mathbb{R}\cup\{-\infty\}$. The* maximum ergodic average *of (f_n) is defined to be the quantity*

$$\beta[(f_n)] := \sup_{\mu \in \mathcal{M}_T} \lim_{n\to\infty} \frac{1}{n} \int f_n\, d\mu = \sup_{\mu \in \mathcal{M}_T} \inf_{n\ge 1} \frac{1}{n} \int f_n\, d\mu.$$

We define $\mathcal{M}_{\max}[(f_n)]$ to be the set of all $\mu \in \mathcal{M}_T$ for which this supremum is attained.

The following important result, called the *semi-uniform subadditive ergodic theorem* in [45], is due independently to S. J. Schreiber [43] and to R. Sturman and J. Stark [45]. Since the version which we use is somewhat more general than those given by Schreiber and Sturman-Stark, we include a proof in the appendix.

Theorem 3.4 (Semi-uniform subadditive ergodic theorem). *Let (f_n) be a subadditive sequence of upper semicontinuous functions from X to $\mathbb{R} \cup \{-\infty\}$. Then*

$$\beta[(f_n)] = \lim_{n \to \infty} \sup_{x \in X} \frac{1}{n} f_n(x) = \sup_{x \in X} \limsup_{n \to \infty} \frac{1}{n} f_n(x) = \lim_{n \to \infty} \sup_{\mu \in \mathcal{M}_T} \frac{1}{n} \int f_n \, d\mu.$$

We next prove some results describing the structure of the set $\mathcal{M}_{\max}[(f_n)]$ for a subadditive sequence (f_n).

Lemma 3.5. *Let (f_n) be a subadditive sequence of upper semi-continuous functions from X to $\mathbb{R} \cup \{-\infty\}$. Then $\mathcal{M}_{\max}[(f_n)]$ is a compact subset of $\mathcal{M}_T$ and contains an ergodic measure.*

Proof. If $\beta[(f_n)] = -\infty$ then $\mathcal{M}_{\max}[(f_n)] = \mathcal{M}_T$ and the result is trivial. We therefore assume $\beta[(f_n)] \in \mathbb{R}$. By Lemma 3.1 each of the maps $\mu \mapsto (1/n) \int f_n \, d\mu$ is upper semi-continuous, and it follows from this that the map $\mu \mapsto \inf_{n \geq 1} (1/n) \int f_n \, d\mu$ is upper semi-continuous also. Since $\mathcal{M}_T$ is compact this implies that $\mathcal{M}_{\max}[(f_n)]$ is compact and nonempty.

Let $\mu \in \mathcal{M}_{\max}[(f_n)]$. By the ergodic decomposition theorem, there exist a measurable space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable function $\mu_{(\cdot)} \colon \Omega \to \mathcal{M}_T$ such that μ_ω is ergodic $\mathbb{P}$-a.e. and such that for each Borel set $A \subseteq X$ the map $\omega \mapsto \mu_\omega(A)$ is $\mathcal{F}$-measurable and satisfies $\mu(A) = \int_\Omega \mu_\omega(A) d\mathbb{P}(\omega)$. For each $r, k \in \mathbb{N}$ define

$$\mathcal{Z}_{r,k} = \left\{ \omega \in \Omega \colon \frac{1}{r} \int f_r \, d\mu_\omega < \beta[(f_n)] - \frac{1}{k} \right\} \in \mathcal{F}.$$

If one has $\mathbb{P}(\mathcal{Z}_{r,k}) > 0$ for some $r, k \in \mathbb{N}$ then

$$\beta[(f_n)] \leq \frac{1}{r} \int f_r \, d\mu = \frac{1}{r} \int_\Omega \int f_r \, d\mu_\omega d\mathbb{P}(\omega) \leq \beta[(f_n)] \left(1 - \frac{\mathbb{P}(\mathcal{Z}_{r,k})}{k} \right) < \beta[(f_n)],$$

a contradiction. We conclude that $\mathbb{P}(\mathcal{Z}_{r,k}) = 0$ for all $r, k \in \mathbb{N}$ and thus

$$\mathbb{P}\left(\{\omega \in \Omega \colon \mu_\omega \in \mathcal{M}_{\max}[(f_n)]\} \right) = \mathbb{P}\left(\left\{ \omega \in \Omega \colon \inf_{r \geq 1} \frac{1}{r} \int f_r \, d\mu_\omega \geq \beta[(f_n)] \right\} \right) = 1.$$

In particular there exists $\omega \in \Omega$ such that μ_ω is ergodic and $\mu_\omega \in \mathcal{M}_{\max}[(f_n)]$. $\quad\square$

The following result gives an analogue of the *subordination principle* described by T. Bousch [7]. While only parts of its statement are actually required in this article, the full statement is included for the sake of interest.

Lemma 3.6. *Let (f_n) be a subadditive sequence of upper semi-continuous functions from X to $\mathbb{R} \cup \{-\infty\}$, and suppose that there exists $\lambda \in \mathbb{R}$ such that $\sup\{f_n(x) \colon x \in X\} = n\lambda$ for all $n \in \mathbb{N}$. Then $\beta[(f_n)] = \lambda$ and if we define for each n*

$$Y_n := \{x \in X \colon f_n(x) = n\lambda\}$$

then $Y := \bigcap_{n=1}^\infty Y_n$ is compact and nonempty and satisfies $TY \subseteq Y$. Furthermore, each $\mu \in \mathcal{M}_T$ satisfies $\mathcal{M}_{\max}[(f_n)]$ if and only if it satisfies $\mu(Y) = 1$.

Proof. Since $\sup f_n = n\lambda$ for each n it is clear that each Y_n is closed and that $\beta[(f_n)] \leq \lambda$. If $x \in Y_{n+1}$ then since

$$(n+1)\lambda = f_{n+1}(x) \leq f_n(x) + f_1(T^n x) \leq f_n(x) + \lambda \leq (n+1)\lambda$$

we have $x \in Y_n$ also. It follows that the intersection $\bigcap_{n=1}^\infty Y_n$ is nonempty. If $x \in Y$ then for each $n \in \mathbb{N}$ we have

$$(n+1)\lambda = f_{n+1}(x) \leq f_n(Tx) + f_1(x) = f_n(Tx) + \lambda \leq (n+1)\lambda$$

so that $f_n(Tx) = n\lambda$, and we deduce that $Tx \in Y$. By the Krylov-Bogolioubov theorem there exists at least one invariant measure μ such that $\mu(Y) = 1$. Since then $n^{-1} \int f_n \, d\mu = \lambda$ for every $n \in \mathbb{N}$ it follows that $\beta[(f_n)] \geq \lambda$, and this argument also shows that if $\nu(Y) = 1$ and $\nu \in \mathcal{M}_T$ then necessarily $\nu \in \mathcal{M}_{\max}[(f_n)]$. Finally, suppose that $\mu \in \mathcal{M}_T$ with $\mu(X \setminus Y) > 0$. Choose $r \in \mathbb{N}$, $\delta > 0$ and a nonempty open set $U \subseteq X \setminus Y$ such that $\mu(U) > 0$ and $f_r(x) < r(\lambda - \delta)$ for all $x \in U$. We have

$$\inf_{n \geq 1} \frac{1}{n} \int f_n \, d\mu \leq \frac{1}{r} \int f_r \, d\mu \leq (1 - \mu(U))\lambda + \mu(U)(\lambda - \delta) < \lambda = \beta[(f_n)]$$

and therefore $\mu \notin \mathcal{M}_{\max}[(f_n)]$. $\qquad\square$

The proposition given below will be needed to make use of the hypothesis $\|\mathcal{A}(x,n)\| \leq M$ in the proof of Theorem 2.1. The proof is not dissimilar to [37, Theorem 1].

Proposition 3.7. *Suppose that $T \colon X \to X$ is minimal. Let (f_n) be a subadditive sequence of upper semi-continuous functions from X to $\mathbb{R} \cup \{-\infty\}$. Suppose that there exists $C \in \mathbb{R}$ such that $f_n(x) \leq C$ for all $n \in \mathbb{N}$ and $x \in X$. Then either $|f_n(x)| \leq C$ for all $n \in \mathbb{N}$ and $x \in X$, or $\lim_{n \to \infty} \frac{1}{n} \sup_{x \in X} f_n(x) < 0$.*

Proof. If $C < 0$ then the result is trivial since $\sup f_n \leq n \sup f_1 \leq nC$ for each $n \in \mathbb{N}$, so we assume $C \geq 0$. Using Theorem 3.4 and Lemma 3.5 we may take $\lambda \in \mathbb{R} \cup \{-\infty\}$ and an ergodic measure $\mu \in \mathcal{M}_T$ such that

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \sup_{x \in X} f_n(x) = \inf_{n \geq 1} \frac{1}{n} \int f_n \, d\mu.$$

Suppose that $f_N(z) < -(C + \varepsilon) < 0$ for some $z \in X$ and $N, \varepsilon > 0$. Using the semi-continuity of f_N, choose a nonempty open set $U \subseteq X$ such that $f_N(x) < -(C + \varepsilon)$ for all $x \in U$. Since T is minimal we have $\mu(U) > 0$.

Using the Birkhoff ergodic theorem and the subadditive ergodic theorem respectively, choose $x_0 \in U$ such that $n^{-1} \sum_{k=0}^{n-1} \chi_U(T^k x_0) \to \mu(U)$ and $n^{-1} f_n(x_0) \to \lambda$. Let $(m_j)_{j=0}^{\infty}$ be the increasing sequence of integers given by $m_0 = 0$ and $m_{j+1} = \min\{m > m_j \colon T^{m_j} x_0 \in U\}$. Now let $(n_r)_{r=0}^{\infty}$ be given by $n_r = m_{Nr}$ so that $n_{r+1} \geq n_r + N$ and $T^{n_r} x_0 \in U$ for each $r \geq 0$. Note that $\lim_{r \to \infty} r/n_r = \mu(U)/N > 0$. For each $r \in \mathbb{N}$ we have

$$f_{n_r}(x_0) \leq \sum_{k=1}^{r} \left(f_N \left(T^{n_{k-1}} x_0 \right) + f_{n_k - n_{k-1} - N} \left(T^{n_{k-1}+N} x_0 \right) \right) \leq -r\left(C + \varepsilon \right) + rC$$

and hence

$$\lim_{n \to \infty} \frac{1}{n} \sup_{x \in X} f_n(x) = \lambda = \lim_{r \to \infty} \frac{1}{n_r} f_{n_r}(x_0) \leq -\frac{\mu(U)\varepsilon}{N} < 0.$$

The proof is complete. $\qquad\square$

4. Proof of Theorem 2.1

We require the following simple result on the metric d_{Gr}.

Lemma 4.1. *Let $V, W \in \mathrm{Gr}(p, d)$ where $1 \leq p \leq d$. Then,*

$$d_{\mathrm{Gr}}(V, W) = \max_{\substack{v \in V \\ \|v\|=1}} \mathrm{dist}(v, W).$$

Proof. Note that $d_{\mathrm{Gr}}(V, W) \leq 1$ for every $V, W \in \mathrm{Gr}(p, d)$, see e.g. [30, p.56]. Let $P_V^\perp$ and $P_W^\perp$ denote the operators of orthogonal projection onto V and W respectively. If $\max\{\mathrm{dist}(v, W) \colon v \in V, \|v\| = 1\} = 1$ then the result is clear. Otherwise, since

$$\|(I - P_W^\perp)P_V^\perp\| = \max_{\|v\|=1} \mathrm{dist}(Pv, W) = \max_{\substack{v \in V \\ \|v\|=1}} \mathrm{dist}(v, W)$$

the result follows from [30, Theorem I-6.34]. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

For each $B \in \mathrm{Mat}_d(\mathbb{C})$ write $|B| := \sqrt{B^*B}$, and for $1 \leq i \leq d$ let $\sigma_1(B) \geq \ldots \geq \sigma_d(B)$ denote the eigenvalues of $|B|$ listed in decreasing order, allowing repetitions if multiplicities occur. Clearly $0 \leq \sigma_i(B) \leq \|B\|$ for every i. The values $\sigma_i(B)$ depend continuously on $B \in \mathrm{Mat}_d(\mathbb{C})$, and if $A, B \in \mathrm{Mat}_d(\mathbb{C})$ then for $1 \leq \ell \leq d$,

$$\prod_{i=1}^\ell \sigma_i(AB) \leq \left(\prod_{i=1}^\ell \sigma_i(A) \right) \left(\prod_{i=1}^\ell \sigma_i(B) \right),$$

see e.g. [20]. For each $x \in X$, $n \in \mathbb{N}$ and $1 \leq \ell \leq d$ let us define $f_n^\ell(x) = \sum_{i=1}^\ell \log \sigma_i(\mathcal{A}(x, n))$. Each (f_n^ℓ) is a subadditive sequence of continuous functions from X to $\mathbb{R} \cup \{-\infty\}$ and the results of §3 may therefore be applied.

Let $M \geq 1$ such that $\|\mathcal{A}(x, n)\| \leq M$ for all $x \in X$ and all $n \in \mathbb{N}$. For each integer ℓ in the range $1 \leq \ell \leq d$, define

$$\theta_\ell := \lim_{n \to \infty} \sup_{x \in X} \frac{1}{n} \sum_{i=1}^\ell \log \sigma_i(\mathcal{A}(x, n))$$

which exists by Theorem 3.4. For $x \in X$, $n \in \mathbb{N}$ and $1 < \ell \leq d$ we have

$$\sum_{i=1}^\ell \log \sigma_i(\mathcal{A}(x, n)) \leq \sum_{i=1}^{\ell-1} \log \sigma_i(\mathcal{A}(x, n)) + \log M \leq \ell \log M$$

and therefore $\theta_{\ell+1} \leq \theta_\ell \leq 0$ for $1 \leq \ell < d$. If $\theta_1 < 0$ then Theorem 2.1 is vacuously true with $p = 0$, $\mathcal{V}(x) \equiv \{0\}$ and $\mathcal{W}(x) \equiv \mathbb{C}^d$, so we henceforth assume $\theta_1 = 0$. Take $p \in \mathbb{N}$ such that $\theta_\ell = 0$ for $1 \leq \ell \leq p$ and $\theta_\ell < 0$ for $p < \ell \leq d$. Applying Proposition 3.7 to (f_n^ℓ) it follows that for $1 \leq \ell \leq p$

$$-\ell \log M \leq \sum_{i=1}^\ell \log \sigma_i(\mathcal{A}(x, n)) \leq \ell \log M$$

for all $x \in X$ and $n \in \mathbb{N}$. We conclude from this that there is $\delta_0 > 0$ such that

$$\min_{1 \leq i \leq p} \inf_{x \in X} \inf_{n \geq 1} \sigma_i(\mathcal{A}(x, n)) \geq \delta_0.$$

Since $\theta_i < 0$ for $p < i \leq d$ we similarly deduce that there exist $C_0 > 0$, $\xi \in (0, 1)$ such that for each $n \in \mathbb{N}$

$$\max_{p < i \leq d} \sup_{x \in X} \sigma_i(\mathcal{A}(x, n)) \leq C_0 \xi^n.$$

Given $x \in X$ and $n \in \mathbb{N}$, let $U_n^+(x) \in \mathrm{Gr}(p, d)$ be the vector space spanned by those eigenvectors of $|\mathcal{A}(x, n)|$ which correspond to the eigenvalues $\sigma_1(\mathcal{A}(x, n))$ up to $\sigma_p(\mathcal{A}(x, n))$ and let $U_n^-(x) \in \mathrm{Gr}(d - p, d)$ be the space spanned by those

eigenvectors associated to the remaining eigenspaces. If v is an eigenvector of $|\mathcal{A}(x,n)|$ with eigenvalue $\sigma_i(A(x,n))$ then

(3)
$$\|\mathcal{A}(x,n)v\|^2 = \langle \mathcal{A}(x,n)v, \mathcal{A}(x,n)v \rangle = \langle \mathcal{A}(x,n)^*\mathcal{A}(x,n)v, v \rangle = \sigma_i(\mathcal{A}(x,n))^2 \|v\|^2.$$

Since $|\mathcal{A}(x,n)|$ is a normal matrix there exists an orthonormal basis for $\mathbb{C}^d$ consisting of its eigenvectors. In particular $U_n^+(x)$ is orthogonal to $U_n^-(x)$, and using (3) we may derive

$$\inf\left\{\|\mathcal{A}(x,n)v\| \colon v \in U_n^+(x) \text{ and } \|v\| = 1\right\} \geq \delta_0,$$

$$\sup\left\{\|\mathcal{A}(x,n)v\| \colon v \in U_n^-(x) \text{ and } \|v\| = 1\right\} \leq C_0\xi^n$$

for all $x \in X$ and $n \in \mathbb{N}$.

We now construct the function $\mathcal{V}$ and establish its properties. The essential idea is to show for each $x \in X$ that the sequence $\mathcal{A}(x,n)U_{2n}^+(x)$ forms a Cauchy sequence in $\mathrm{Gr}(p,d)$ and to define $\mathcal{V}(x)$ to be its limit. This is related to the construction in [19], but our argument is simplified by the presence of estimates which are uniform in x.

Let $x \in X$, $n \in \mathbb{N}$ and $v \in \mathbb{C}^d$; if $\|\mathcal{A}(x,n)v\| \geq \varepsilon\|v\|$ for some $\varepsilon > 0$, an easy calculation shows that for $1 \leq k < n$

(4)
$$\|\mathcal{A}(x,k)v\| \geq M^{-1}\varepsilon\|v\|$$

and

(5)
$$\|\mathcal{A}(T^kx, n-k)\mathcal{A}(x,k)v\| = \|\mathcal{A}(x,n)v\| \geq M^{-1}\varepsilon\|\mathcal{A}(x,k)v\|.$$

For each $n \geq 1$, $\kappa > 0$ and $x \in X$, define a subset of $\mathrm{Gr}(p,d)$ by

$$\mathfrak{V}(x,n,\kappa) := \left\{\mathcal{A}(T^{-n}x,n)W \colon W \in \mathrm{Gr}(p,d), \|\mathcal{A}(T^{-n}x,2n)w\| \geq \kappa\|w\| \, \forall\, w \in W\right\}.$$

Note that for $\kappa \leq \delta_0$ we have $\mathcal{A}(T^{-n}x,n)U_{2n}^+(T^{-n}x) \in \mathfrak{V}(x,n,\kappa)$ and so the latter set is nonempty. Moreover we have

(6)
$$\mathfrak{V}(x,n,\kappa) \subseteq \mathfrak{V}\left(x,k,M^{-2}\kappa\right)$$

for $x \in X$ and $1 \leq k \leq n$ and

(7)
$$\mathcal{A}(x)\mathfrak{V}(x,n,\kappa) \subseteq \mathfrak{V}\left(x,n-1,M^{-1}\kappa\right)$$

by virtue of (4) and (5). We claim that for each $n \in \mathbb{N}$ and $x \in X$,

(8)
$$\mathrm{diam} \bigcup_{r=n}^{\infty} \mathfrak{V}(x,r,\kappa) \leq \kappa^{-1}C_1\xi^n$$

where $C_1 := 2C_0M$. Suppose that

$$\mathcal{A}\left(T^{-(n+m)}x, n+m\right)w \in \mathcal{A}\left(T^{-(n+m)}x, n+m\right)W \in \mathfrak{V}(x,n+m,\kappa).$$

Let P be given by orthogonal projection from $\mathbb{C}^d$ onto $U_n^+(T^{-n}x)$. We have

$$\left\|\mathcal{A}\left(T^{-(n+m)}x, n+m\right)w - \mathcal{A}\left(T^{-n}x, n\right)P\mathcal{A}\left(T^{-(n+m)}x, m\right)w\right\|$$

$$\leq C_0\xi^n\left\|\mathcal{A}\left(T^{-(n+m)}x, m\right)w\right\| \leq C_0M\xi^n\left\|\mathcal{A}\left(T^{-(n+m)}x, n+m\right)w\right\|$$

where we have used (5). It follows that

$$\mathrm{dist}(v, \mathcal{A}(T^{-n}x,n)U_n^+(x)) \leq MC_0\kappa^{-1}\xi^n\|v\|$$

for all $v \in \mathcal{A}\left(T^{-(n+m)}x, n+m\right)W$ and therefore (8) holds by Lemma 4.1. We deduce that for each $x \in X$ the set

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{r=n}^{\infty} \mathfrak{V}(x, r, \kappa)}$$

contains a unique element for each $\kappa \leq \delta_0$. Since clearly $\mathfrak{V}(x, n, \kappa_1) \subseteq \mathfrak{V}(x, n, \kappa_2)$ for $\kappa_2 \leq \kappa_1$ it follows that this element does not depend on κ. Denote this element by $\mathcal{V}(x)$. We have $\mathcal{A}(x)\mathcal{V}(x) = \mathcal{V}(Tx)$ as an easy consequence of (7). Now take n large enough that $\delta_0^{-1}C_1\xi^n < \delta_0/3M$ and let P be given by orthogonal projection onto some arbitrarily selected element of $\mathfrak{V}(x, n, \delta_0)$. Given any $v \in \mathcal{V}(x)$, we have $\|v - Pv\| \leq (\delta_0/3M)\|v\|$ as a consequence of (8). In particular this implies $\|Pv\| \geq (2/3)\|v\|$. We have

$$\|\mathcal{A}(x, n)v\| \geq \|\mathcal{A}(x, n)Pv\| - \|\mathcal{A}(x, n)v - \mathcal{A}(x, n)Pv\|$$

$$\geq \delta_0\|Pv\| - M\|v - Pv\| \geq (\delta_0/3)\|v\|.$$

It follows from (4) that for all $x \in X$ and every $n \in \mathbb{N}$ we have $\|\mathcal{A}(x, n)v\| \geq (\delta_0/3M)\|v\|$ for every $v \in \mathcal{V}(x)$.

It remains to show that $\mathcal{V}(x)$ depends continuously on x. Define $\delta := \delta_0/3M$. Let $n \in \mathbb{N}$ and suppose that $x, y \in X$ satisfy

$$\max\left\{\|\mathcal{A}(T^{-n}x, 2n) - \mathcal{A}(T^{-n}y, 2n)\|, \|\mathcal{A}(T^{-n}x, n) - \mathcal{A}(T^{-n}y, n)\|\right\} \leq \delta\xi^n.$$

If $w \in \mathcal{V}(T^{-n}x)$ then

$$\|\mathcal{A}(T^{-n}y, 2n)w\| \geq \|\mathcal{A}(T^{-n}x, 2n)w\| - \delta\xi^n\|w\| \geq (1 - \xi)\delta\|w\|$$

and it follows that $\mathcal{A}(T^{-n}y, n)\mathcal{V}(T^{-n}x) \in \mathfrak{V}(y, n, (1 - \xi)\delta)$. If $v = \mathcal{A}(T^{-n}x, n)w \in \mathcal{V}(x) = \mathcal{A}(T^{-n}x, n)\mathcal{V}(T^{-n}x)$, then

$$\|\mathcal{A}(T^{-n}x, n)w - \mathcal{A}(T^{-n}y, n)w\| \leq \delta\xi^n\|w\| \leq \xi^n\|v\|.$$

It follows from Lemma 4.1 that $d_{\mathrm{Gr}}(\mathcal{V}(x), \mathcal{A}(T^{-n}y, n)\mathcal{V}(T^{-n}x)) \leq \xi^n$, and therefore $d_{\mathrm{Gr}}(\mathcal{V}(x), \mathcal{V}(y)) \leq (1 + C_1(1 - \xi)^{-1}\delta^{-1})\xi^n$ as required.

We next construct the function $\mathcal{W}$ and establish its properties. Similarly to the case of $\mathcal{V}$, the idea is to show that $U_n^-(x)$ forms a Cauchy sequence and to define $\mathcal{W}(x)$ to be its limit. This section of the proof thus more closely approaches certain proofs of the Oseledec multiplicative ergodic theorem such as that given in [42], though as before we differ from the measurable case in that we require uniform estimates.

For each $n \in \mathbb{N}$, $x \in X$ and $K > 0$ define

$$\mathfrak{W}(x, n, K) = \{W \in \mathrm{Gr}(d - p, d) \colon \|\mathcal{A}(x, n)v\| \leq K\xi^n\|v\| \text{ for all } v \in W\}.$$

Note that $U_n^-(x) \in \mathfrak{W}(x, n, K)$ for every $K \geq C_0$ and in particular $\mathfrak{W}(x, n, K)$ is nonempty. We assert that for each $n \in \mathbb{N}$ we have

$$(9) \qquad \qquad \mathrm{diam}\, \bigcup_{r=n}^{\infty} \mathfrak{W}(x, r, K) \leq KC_2\xi^n$$

where $C_2 = 2\delta_0^{-1}(M + 1)$. Suppose that $r \geq n \in \mathbb{N}$ and $W \in \mathfrak{W}(x, r, K)$ are given, and let $v \in W$. Write $v = u_1 + u_2$ with $u_1 \in U_n^+(x)$ and $u_2 \in U_n^-(x)$; we have $\|u_2\| \leq \|v\|$ since the two spaces are mutually orthogonal. Since

$$\|\mathcal{A}(x, r)(v - u_2)\| \leq K\xi^r\|v\| + MC_0\xi^n\|u_2\| \leq (MC_0 + K)\xi^n\|v\|$$

and
$$\|\mathcal{A}(x,r)(v-u_2)\| = \|\mathcal{A}(x,r)u_1\| \geq \delta_0\|u_1\|$$

we have
$$\operatorname{dist}(v, U_n^-(x)) = \|u_1\| \leq \delta_0^{-1}(MC_0 + K)\xi^n\|v\| \leq \delta_0^{-1}(M+1)K\xi^n\|v\|$$

and (9) holds by Lemma 4.1. It follows that for each $x \in X$ and $K \geq C_0$, the set
$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{r=n}^{\infty} \mathfrak{W}(x,r,K)}$$

contains a unique element, which we denote by $\mathcal{W}(x)$. Since clearly $\mathfrak{W}(x,n,K_1) \subseteq \mathfrak{W}(x,n,K_2)$ when $K_1 \leq K_2$ the definition of $\mathcal{W}(x)$ is not influenced by the choice of $K \geq C_0$. Given $x \in X$ and $n \in \mathbb{N}$, let P be given by orthogonal projection onto $U_n^-(x) \in \mathfrak{W}(x,n,C_0)$. For each $w \in \mathcal{W}(x)$ we have $\|w - Pw\| \leq C_0C_2\xi^n\|w\|$ as a consequence of (9) and hence
$$\|\mathcal{A}(x,n)w\| \leq \|\mathcal{A}(x,n)(w-Pw)\| + \|\mathcal{A}(x,n)Pw\|$$
$$\leq MC_0C_2\xi^n\|w\| + C_0\xi^n\|Pw\| \leq (MC_2+1)C_0\xi^n\|w\|$$

as required for the statement of Theorem 2.1.

We next prove that $\mathcal{W}(x)$ depends continuously on x. Let $x, y \in X$ and suppose that $d(x,y)$ is small enough that $\|\mathcal{A}(x,n) - \mathcal{A}(y,n)\| \leq \xi^n$. Since for any $w \in \mathcal{W}(x)$,
$$\|\mathcal{A}(y,n)w\| \leq \xi^n\|w\| + \|\mathcal{A}(x,n)w\| \leq (MC_2C_0 + C_0 + 1)\xi^n\|w\|,$$

we have $\mathcal{W}(x) \in \mathfrak{W}(y,n,MC_2C_0 + C_0 + 1)$ and it follows from (9) that
$$d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y)) \leq (MC_2C_0 + C_0 + 1)C_2\xi^n.$$

The following standard argument shows that $\mathcal{W}$ is invariant. For each $x \in X$ define
$$\tilde{\mathcal{W}}(x) = \left\{ v \in \mathbb{C}^d \colon \limsup_{n \to \infty} \|A_x^n v\|^{1/n} < 1 \right\}.$$

Clearly $\tilde{\mathcal{W}}(x)$ is a linear subspace of $\mathbb{C}^d$, $\mathcal{W}(x) \subseteq \tilde{\mathcal{W}}(x)$, and $\mathcal{A}(x)\tilde{\mathcal{W}}(x) \subseteq \tilde{\mathcal{W}}(Tx)$. If $\dim \tilde{\mathcal{W}}(x) > \dim \mathcal{W}(x)$ then $\tilde{\mathcal{W}}(x) \cap \mathcal{V}(x) \neq \{0\}$ which clearly entails a contradiction. It follows that $\tilde{\mathcal{W}}(x) = \mathcal{W}(x)$ for all $x \in Z$ and therefore $\mathcal{A}(x)\mathcal{W}(x) = \mathcal{A}(x)\tilde{\mathcal{W}}(x) \subseteq \tilde{\mathcal{W}}(Tx) = \mathcal{W}(Tx)$, which concludes our study of the properties of $\mathcal{W}$.

For each $x \in X$ let $P(x)$ denote the projection having image $\mathcal{V}(x)$ and kernel $\mathcal{W}(x)$. It remains to prove that $P(x)$ depends continuously on x. We will show that for every $x \in X$, if y satisfies

(10) $$3\|P(x)\|.[d_{\mathrm{Gr}}(\mathcal{V}(x), \mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y))] < \frac{1}{2}$$

then

(11) $$\|P(x) - P(y)\| \leq 12\|P(x)\|.[d_{\mathrm{Gr}}(\mathcal{V}(x), \mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y))].$$

Since X is compact we may deduce that $\sup \|P\|$ is finite and the result follows.

For notational convenience we write $Q(x) = I - P(x)$ for all $x \in X$. Fix $x \in X$, and for each $y \in X$ define $U(x,y) = P_{\mathcal{V}(y)}^{\perp}P(x) + P_{\mathcal{W}(y)}^{\perp}Q(x)$, where $P_Z^{\perp}$ denotes orthogonal projection onto Z. Since $I = P(x) + Q(x) = P_{\mathcal{V}(x)}^{\perp}P(x) + P_{\mathcal{W}(x)}^{\perp}Q(x)$ we have

(12) $$\|U(x,y) - I\| \leq (2\|P(x)\| + 1).[d_{\mathrm{Gr}}(\mathcal{V}(x), \mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y))].$$

Suppose that y satisfies (10). Then $U(x,y)$ is invertible and

$$(13) \qquad \|U(x,y)^{-1} - I\| \leq \sum_{n=1}^{\infty} \|U(x,y) - I\|^n$$

$$\leq 6\|P(x)\| \left[d_{\mathrm{Gr}}(\mathcal{V}(x), \mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y)) \right].$$

Since for each $v \in \mathcal{V}(x)$ and $w \in \mathcal{W}(x)$ we have

$$U(x,y)P(x)(v+w) = U(x,y)v = P(y)U(x,y)(v+w)$$

it follows that $P(y) = U(x,y)P(x)U(x,y)^{-1}$. Combining this with (12) and (13) yields (11) and the proof is complete.

5. Proof of Theorem 2.2

Let $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ be compact and product bounded with $\varrho(\mathsf{A}) = 1$, let $\|\!|\cdot|\!\|$ be an extremal norm for A, and choose $M > 0$ such that $\|v\| \leq M\|\!|v|\!\| \leq M^2\|v\|$ for all $v \in \mathbb{C}^d$. As in the introduction we let $\mathcal{A} \colon \mathsf{A}^{\mathbb{Z}} \to \mathrm{Mat}_d(\mathbb{C})$ be given by projection onto the zeroth co-ordinate, let $T \colon \mathsf{A}^{\mathbb{Z}} \to \mathsf{A}^{\mathbb{Z}}$ be given by the shift map, and take d to be the metric on $\mathsf{A}^{\mathbb{Z}}$ defined previously. Clearly $\mathcal{A}$ and T are Lipschitz continuous. For each $n \in \mathbb{N}$ we have

$$\left\{ \mathcal{A}(x,n) \colon x \in \mathsf{A}^{\mathbb{Z}} \right\} = \{ A_n \cdots A_1 \colon A_i \in \mathsf{A} \}$$

and therefore $\sup \left\{ \log \|\!|\mathcal{A}(x,n)|\!\| \colon x \in \mathsf{A}^{\mathbb{Z}} \right\} = 0$ for all $n \in \mathbb{N}$. By Lemma 3.6 the set

$$Y := \{ x \colon \|\!|\mathcal{A}(x,n)|\!\| = 1 \text{ for all } n \geq 1 \}$$

is compact and nonempty and satisfies $TY \subseteq Y$.

Let $Z = TZ$ be any minimal set contained in Y. Note that for all $x \in Z$ and $n \in \mathbb{N}$ we have $\|\mathcal{A}(x,n)\| \leq M^2$ since $\|\!|\mathcal{A}(x,n)|\!\| = 1$. We may therefore apply Theorem 2.1 to the minimal set Z and the cocycle $\mathcal{A}$. If $p = 0$ then we would have $\|\!|\mathcal{A}(x,n)|\!\| < 1$ for some $x \in Z$ and $n \in \mathbb{N}$, so it must be the case that $p \in \mathbb{N}$. To prove Theorem 2.2, we must show firstly that the functions $\mathcal{V}, \mathcal{W}$ and P provided by Theorem 2.1 are Hölder continuous, and secondly that for all $x \in Z$ and $n \in \mathbb{N}$ one has $\|\!|\mathcal{A}(x,n)v|\!\| = \|\!|v|\!\|$ for every $v \in \mathcal{V}(x)$.

The proof of Hölder continuity is straightforward. Let δ, ξ be as given by Theorem 2.1. Given any $\varepsilon > 0$, choose $C_\varepsilon > 0$ such that $2C_\varepsilon n e^{-n\varepsilon} \leq 1$ for all $n \in \mathbb{N}$. If $d(x,y) \leq C_\varepsilon \delta M^{-4} e^{-n\varepsilon} \xi^n 2^{-n}$ then

$$\max_{-n \leq k \leq n} \left\| \mathcal{A}(T^k x) - \mathcal{A}(T^k y) \right\| \leq 2^n \sum_{i \in \mathbb{Z}} \frac{\left\| \mathcal{A}(T^i x) - \mathcal{A}(T^i y) \right\|}{2^{|i|}} \leq C_\varepsilon \delta M^{-4} \xi^n e^{-n\varepsilon}$$

and therefore

$$\left\| \mathcal{A}(T^{-n} x, 2n) - \mathcal{A}(T^{-n} y, 2n) \right\|$$

$$\leq \sum_{i=-n}^{n-1} \left\| \mathcal{A}(T^{i+1} x, n-i-1) \right\| \cdot \left\| \mathcal{A}(T^i x) - \mathcal{A}(T^i y) \right\| \cdot \left\| \mathcal{A}(T^{-n} y, n+i) \right\|$$

$$\leq 2C_\varepsilon n \delta \xi^n e^{-\varepsilon n} \leq \delta \xi^n$$

where we adopt the convention $\mathcal{A}(\cdot, 0) \equiv I$. The same estimate clearly also yields $\|\mathcal{A}(T^{-n}x, n) - \mathcal{A}(T^{-n}y, n)\| \leq \delta \xi^n$ and $\|\mathcal{A}(x,n) - \mathcal{A}(y,n)\| \leq \delta \xi^n \leq \xi^n$. Applying Theorem 2.1 we deduce that $d_{\mathrm{Gr}}(\mathcal{V}(x), \mathcal{V}(y))$ and $d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y))$ are both

bounded by $\tilde{C}\xi^n$. It follows that for $\alpha := \log\xi/(\log\xi - \log 2 - \varepsilon) > 0$,

$$\sup_{\substack{x,y\in Z\\x\neq y}} \frac{d_{\mathrm{Gr}}(\mathcal{V}(x),\mathcal{V}(y))}{d(x,y)^\alpha} < \infty$$

and similarly for $\mathcal{W}$ so that $\mathcal{V}$ and $\mathcal{W}$ are both α-Hölder continuous. By Theorem 2.1 this implies that P is α-Hölder continuous also.

We now prove that for every $x \in Z$ and $n \in \mathbb{N}$ we have $\|\mathcal{A}(x,n)v\| = \|v\|$ for every $v \in \mathcal{V}(x)$. For each $x \in Z$ define

$$\mathcal{S}(x) := \left\{ B\colon \liminf_{n\to\infty} \max\left[d(T^n x, x), \|\mathcal{A}(x,n) - B\|\right] = 0 \right\}.$$

Since T acts minimally on Z, x is recurrent, and since $\|\mathcal{A}(x,n)\| = 1$ for each n the set $\mathcal{S}(x)$ is nonempty. If $\lim_{k\to\infty} B_k = B$ with each $B_k \in \mathcal{S}(x)$ then we may choose a strictly increasing sequence (n_k) such that $d(T^{n_k}x, x) < 1/k$, $\|B_k - B\| \leq 1/k$ and $\|\mathcal{A}(x,n_k) - B_k\| < 1/k$ for each $k \in \mathbb{N}$, which shows that $B \in \mathcal{S}(x)$ and therefore $\mathcal{S}(x)$ is closed. Since clearly $\|B\| = 1$ for all $B \in \mathcal{S}(x)$ it follows that $\mathcal{S}(x)$ is compact.

We claim that $\mathcal{S}(x)$ is a semigroup. Let $B_1, B_2 \in \mathcal{S}(x)$; it suffices to show that for any $N, \varepsilon > 0$ there is $n > N$ such that $d(T^n x, x) < \varepsilon$ and $\|\mathcal{A}(x,n) - B_1 B_2\| < \varepsilon$. Since $B_1 \in \mathcal{S}(x)$ we can choose $n_1 > N$ such that $\|\mathcal{A}(x,n_1) - B_1\| < \varepsilon/3$ and $d(T^{n_1}x, x) < \varepsilon/2$. Since $B_2 \in \mathcal{S}(x)$ we may choose $n_2 > N$ such that $\|\mathcal{A}(x,n_2) - B_2\| < \varepsilon/3$ and such that $d(T^{n_2}x, x)$ is so small as to guarantee $\|\mathcal{A}(T^{n_2}x, n_1) - \mathcal{A}(x,n_1)\| < \varepsilon/3$ and $d(T^{n_1+n_2}x, T^{n_1}x) < \varepsilon/2$. We have

$$\|\mathcal{A}(x, n_1+n_2) - B_1 B_2\| \leq \|\mathcal{A}(T^{n_2}x, n_1)\mathcal{A}(x,n_2) - \mathcal{A}(x,n_1)\mathcal{A}(x,n_2)\|$$
$$+ \|\mathcal{A}(x,n_1)\mathcal{A}(x,n_2) - \mathcal{A}(x,n_1)B_2\|$$
$$+ \|\mathcal{A}(x,n_1)B_2 - B_1 B_2\| < \varepsilon$$

and

$$d(T^{n_1+n_2}x, x) \leq d(T^{n_1+n_2}x, T^{n_1}x) + d(T^{n_1}x, x) < \varepsilon$$

as required to prove the claim.

Given any $B \in \mathcal{S}(x)$, take $(n_r)_{r=1}^\infty$ such that $\mathcal{A}(x, n_r) \to B$ and $d(T^{n_r}x, x) \to 0$. If v is a nonzero element of $\mathcal{V}(x)$ then clearly $\mathcal{A}(x,n_r)v \to Bv$. Since Theorem 2.1 gives $\|\mathcal{A}(x,n_r)v\| \geq \delta\|v\|$ for all $r \in \mathbb{N}$ we have $\|Bv\| \geq \delta\|v\| > 0$. Since $\mathcal{A}(x,n_r)\mathcal{V}(x) = \mathcal{V}(T^{n_r}x)$, $T^{n_r}x \to x$ and $\mathcal{V}$ is continuous it follows that in fact Bv is a nonzero element of $\mathcal{V}(x)$. By a simpler version of the same argument we see that $Bw = 0$ for every $w \in \mathcal{W}(x)$, and we conclude that the image of B is precisely $\mathcal{V}(x)$ whilst the kernel of B is precisely $\mathcal{W}(x)$.

We now finish the proof. Since $\mathcal{S}(x)$ is a compact semigroup, it contains an idempotent element P (see e.g. [27]). If $\|\mathcal{A}(x,k)v\| \leq (1-\varepsilon)\|v\|$ for some vector $v \in \mathcal{V}(x)$ and positive integer k, then $\|\mathcal{A}(x,n)v\| \leq (1-\varepsilon)\|v\|$ for all large enough n and therefore $\|Pv\| \leq (1-\varepsilon)\|v\|$. But since v lies in the image of P we have $v = Pw = P^2w = Pv$ for some $w \in \mathbb{C}^d$, and we conclude that $\|v\|$ must equal zero. It follows that for each $x \in X$ and $n \in \mathbb{N}$ we have $\|\mathcal{A}(x,n)v\| = \|v\|$ for all $v \in \mathcal{V}(x)$ and the theorem is proved.

Remark 1. *Since we have identified both the image and the kernel of the idempotent P, it follows that for each x the semigroup $\mathcal{S}(x)$ in fact contains a* unique *idempotent element, namely the projection $P(x)$. The family of semigroups $\mathcal{S}(x)$ should be compared to the "limit semigroup" introduced by F. Wirth [48].*

6. PROOF OF THEOREM 1.2

The following lemma allows us to ignore cases in which A fails to be product bounded. Results of this kind are used in the proofs of Theorem 1.1 given by Berger-Wang [2], Elsner [18], and Shih *et al.* [44].

Lemma 6.1. *Let $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ be bounded set such that $\varrho(\mathsf{A}) = 1$ and A is not product bounded. Then exist a positive integer $d' < d$ and $U \in GL_d(\mathbb{C})$ such that if P denotes the natural projection from $\mathbb{C}^d$ to $\mathbb{C}^{d'}$ then the set $\hat{\mathsf{A}} := PU^{-1}\mathsf{A}U$ satisfies $\varrho(\mathsf{A}) = 1$, is product bounded and satisfies $\varrho_n^-(\mathsf{A}) \geq \varrho_n^-(\hat{\mathsf{A}})$ for each $n \in \mathbb{N}$.*

Proof. Using [18, Lemma 4] we can find $\hat{\mathsf{A}} = PU^{-1}\mathsf{A}U$ which satisfies all of the required properties except possibly for product boundedness. By repeating this procedure we either obtain a product bounded $\hat{\mathsf{A}}$ with $d' > 1$, or reduce to the case $d = 1$ in which case product boundedness is satisfied automatically. $\square$

If $\varrho(\mathsf{A}) = 0$ then we have nothing to prove, and if $\varrho(\mathsf{A}) > 0$ then by normalising A if necessary we may assume that $\varrho(\mathsf{A}) = 1$. To prove Theorem 1.2, therefore, it suffices by Lemma 6.1 to assume that A is a finite set of $d \times d$ matrices such that $\varrho(\mathsf{A}) = 1$ and A is product bounded. Since A is finite, the metric described in the introduction is Lipschitz equivalent to the more easily-used metric given by

$$d\left[(A_i)_{i \in \mathbb{Z}}, (B_i)_{i \in \mathbb{Z}}\right] = 2^{-\sup\{n \geq 0 \colon A_i = B_i \text{ for } |i| \leq n\}}.$$

The following proposition may be obtained easily by modifying a result of X. Bressaud and A. Quas [11, Theorem 1].

Proposition 6.2. *Let A be finite, let $Z \subseteq \mathsf{A}^{\mathbb{Z}}$ be compact with $TZ = Z$, and let $N \in \mathbb{N}$. Then there exist sequences of integers (r_n), (m_n) and a sequence of points $x_n \in \mathsf{A}^{\mathbb{Z}}$ such that $m_n^{-1} \log n \to 0$ and such that for all sufficiently large n each r_n is divisible by N, $r_n \leq n$, $T^{r_n}x_n = x_n$ and*

$$\max_{0 \leq k < r_n} d(T^k x_n, Z) \leq 2^{-m_n}.$$

Now let $\|\cdot\|$ be an extremal norm for A, let Y be as in Theorem 2.2, and let $Z \subseteq Y$ be any minimal set. Let $\mathcal{V}$, $\mathcal{W}$, P, C and ξ be as given by Theorem 2.2, and define $Q(x) = I - P(x)$ for each $x \in Z$. Note that for $v \in \mathcal{V}(x)$ and $w \in \mathcal{W}(x)$ we have

$$\mathcal{A}(x,n)P(x)(v+w) = \mathcal{A}(x,n)v = P(T^n x)\mathcal{A}(x,n)(v+w)$$

and therefore $\mathcal{A}(x,n)P(x) = P(T^n x)\mathcal{A}(x,n)$ for all $x \in Z$ and $n \in \mathbb{N}$. Clearly this implies that $\mathcal{A}(x,n)Q(x) = Q(T^n x)\mathcal{A}(x,n)$ for all $x \in Z$ and $n \in \mathbb{N}$.

The following two lemmas, and the general strategy of their application, are suggested by [29]. For each $x \in Z$ and $\theta > 0$ let us define

$$\mathfrak{C}(x,\theta) = \left\{v \in \mathbb{C}^d \colon \theta\|P(x)v\| \geq \|Q(x)v\|\right\}.$$

Lemma 6.3. *Let $x, y \in Z$ and suppose that $\|P(x) - P(y)\| \leq \theta < 1/5$. Then $\mathfrak{C}(x,\theta) \subseteq \mathfrak{C}(y, 3\theta)$.*

Proof. If $v \notin \mathfrak{C}(y, 3\theta)$ then $\|Q(y)v\| > 3\theta\|P(y)v\|$ and therefore

$$3\theta\|P(x)v\| \leq 3\theta\|P(y)\| + 3\theta^2\|v\| < \|Q(y)v\| + 3\theta^2\|v\| \leq \|Q(x)v\| + (\theta + 3\theta^2)\|v\|$$

$$\leq (1 + \theta + 3\theta^2)\|Q(x)v\| + (\theta + 3\theta^2)\|P(x)v\|$$

and therefore
$$\theta\|P(x)v\| \leq \frac{2\theta - 3\theta^2}{1 + \theta + 3\theta^2}\|P(x)v\| < \|Q(x)v\|$$
so that $v \notin \mathfrak{C}(x, \theta)$. $\qquad\square$

Lemma 6.4. *Let $x \in Z$ and $n \in \mathbb{N}$, and suppose that $v \in \mathfrak{C}(x, \theta)$ for some $\theta \in (0, 1]$. Then $\mathcal{A}(x, n)v \in \mathfrak{C}(T^n x, K_1\xi^n\theta)$ and $\|\mathcal{A}(x, n)v\| \geq (1 - \theta - K_1\xi^n\theta)\|v\|$, where $K_1 > 0$ does not depend on x, n, θ or v.*

Proof. Let $M = \sup_{z \in Z}\|Q(x)\|$ and $K_1 = 2CM$. If $v \in \mathfrak{C}(x, \theta)$ then clearly
$$\|v\| \leq \|P(x)v\| + \|Q(x)v\| \leq (1 + \theta)\|P(x)v\|.$$

Using Theorem 2.2 it follows that
$$\|P(T^n x)\mathcal{A}(x, n)v\| = \|\mathcal{A}(x, n)P(x)v\| = \|P(x)v\| \geq (1 + \theta)^{-1}\|v\|$$
and
$$\|Q(T^n x)\mathcal{A}(x, n)v\| = \|\mathcal{A}(x, n)Q(x)v\| \leq C_1\xi^n\|Q(x)v\| \leq CM\theta\xi^n\|v\|.$$

Consequently
$$\|\mathcal{A}(x, n)v\| \geq \|P(T^n x)\mathcal{A}(x, n)v\| - \|Q(T^n x)\mathcal{A}(x, n)v\| \geq (1 - \theta - K_1\theta\xi^n)\|v\|$$
and
$$\|Q(T^n x)\mathcal{A}(x, n)v\| \leq K_1\xi^n\theta\|P(T^n x)\mathcal{A}(x, n)v\|$$
as required. $\qquad\square$

We now prove Theorem 1.2. Let $K_2, \alpha > 0$ such that $\|P(x) - P(y)\| \leq K_2 d(x, y)^\alpha$ for all $x, y \in Z$, let $N \geq 1$ be large enough that $K_1\xi^N < 1/3$, and let $(x_n), (m_n), (r_n)$ be as given by Proposition 6.2. Suppose that n is large enough that $K_2 2^{\alpha(N-m_n)} < 1/5$, $m_n \geq N$, and all of the properties listed in Proposition 6.2 are satisfied. Let $q = r_n/N$ and choose $z_1, \ldots, z_q$ such that $d(z_i, T^{(i-1)N}x_n) \leq 2^{-m_n}$ for each i. We then have
$$d(T^N z_i, z_{i+1}) = \max\{d(T^N z_1, T^{iN}x_n), d(T^{iN}x_n, z_{i+1})\} \leq 2^{N-m_n}$$
for $1 \leq i < q$, and similarly $d(T^N z_q, z_1) \leq 2^{N-m_n}$. If $v \in \mathfrak{C}(z_i, K_2 2^{\alpha(N-m_n)})$ for $1 \leq i < q$ then we may apply Lemmas 6.3 and 6.4 to deduce that $\mathcal{A}(z_i, N)v \in \mathfrak{C}(z_{i+1}, K_2 2^{\alpha(N-m_n)})$ and $\|\mathcal{A}(z_i, N)v\| \geq (1 - K_2 2^{1+\alpha(N-m_n)})\|v\|$, and similarly if $v \in \mathfrak{C}(z_q, K_2 2^{\alpha(N-m_n)})$ then $\mathcal{A}(z_q, N)v \in \mathfrak{C}(z_1, K_2 2^{\alpha(N-m_n)})$ and $\|\mathcal{A}(z_q, N)v\| \geq (1 - K_2 2^{1+\alpha(N-m_n)})\|v\|$. It follows that if $v \in \mathfrak{C}(z_1, K_2 2^{\alpha(N-m_n)})$ then
$$\mathcal{A}(x_n, r_n)v = \mathcal{A}(z_q, N) \cdots \mathcal{A}(z_1, N)v \in \mathfrak{C}(z_1, K_2 2^{\alpha(N-m_n)})$$
(where we have used $m_n \geq N$) and
$$\|\mathcal{A}(x_n, r_n)v\| = \|\mathcal{A}(z_q, N) \cdots \mathcal{A}(z_1, N)v\| \geq (1 - K_2 2^{1+\alpha(N-m_n)})^{r_n/N}\|v\|.$$

If we choose $v \in \mathfrak{C}(z_1, K_2 2^{\alpha(N-m_n)})$ with $\|v\| = 1$, then since $r_n \leq n$ we deduce
$$\max_{1 \leq k \leq n} \varrho_k^-(\mathsf{A}) \geq \rho(\mathcal{A}(x_n, r_n))^{1/r_n} = \left(\lim_{k \to \infty}\|\mathcal{A}(x_n, r_n)^k\|^{1/k}\right)^{1/r_n}$$
$$\geq \left(\liminf_{k \to \infty}\|\mathcal{A}(x_n, r_n)^k v\|^{1/k}\right)^{1/r_n}$$
$$\geq (1 - K_2 2^{1+\alpha(N-m_n)})^{1/N} \geq 1 - K_2 2^{1+\alpha(N-m_n)}.$$

It follows that for all large enough n

$$\left| \varrho(\mathsf{A}) - \max_{1 \leq k \leq n} \varrho_k^-(\mathsf{A}) \right| \leq \left(K_2 2^{1+\alpha N} \right) 2^{-\alpha m_n}.$$

To complete the proof we have only to observe that the condition $m_n^{-1} \log n \to 0$ is equivalent to the assertion that $e^{-\varepsilon m_n} = O(1/n^r)$ for every $r, \varepsilon > 0$.

7. Discussion on possible extensions of Theorem 1.2

We shall now briefly discuss some of the limitations of the method of proof of Theorem 1.2 and the prospects for an extension of that theorem using the approach of the present article.

Fix some compact set $\Omega \subset \mathbb{C}^d$, and consider the metric space $\Omega^{\mathbb{Z}}$ equipped with the metric $d[(x_i), (y_i)] = \sum_{i \in \mathbb{Z}} 2^{-|i|} \|x_i - y_i\|$ together with the shift map $T \colon \Omega^{\mathbb{Z}} \to \Omega^{\mathbb{Z}}$. Given a compact T-invariant set $Z \subseteq \Omega^{\mathbb{Z}}$, let us define

$$\varepsilon(Z, n) = \min_{1 \leq k \leq n} \inf_{T^k x = x} \max_{0 \leq i < k} \operatorname{dist}(T^i x, Z).$$

The magnitude of the error term in the proof of Theorem 1.2 is determined by the result of X. Bressaud and A. Quas in [11] which asserts that if Ω is a finite set, then $\varepsilon(Z, n) = O(1/n^r)$ for every $r \in \mathbb{N}$. (To simplify our proof we in fact considered only approximations using periodic orbits whose period is divisible by N, but this requirement could be dispensed with without difficulty.) Bressaud and Quas' result is essentially sharp: see [11] and related work in [12]. In the case where Ω is compact but not finite, the rate of decrease of $\varepsilon(Z, n)$ can be much slower, and this is the principal obstacle in extending Theorem 1.2 to the case in which A compact but infinite. The following simple example illustrates the problem.

Suppose that $\Omega = S^1 \subset \mathbb{C}$. Let $\gamma = (1 - \sqrt{5})/2$ and define

$$Z = \left\{ \left(e^{2\pi i m \gamma} \omega \right)_{m \in \mathbb{Z}} : \omega \in S^1 \right\},$$

which is clearly compact and T-invariant. Let $n \in \mathbb{N}$ and $1 \leq k \leq n$, and suppose that $x \in \Omega^{\mathbb{Z}}$ has $T^k x = x$ and $\max_{0 \leq j < k} \operatorname{dist}(T^j x, Z) \leq 2\varepsilon(Z, n)$. For $j = 0, \ldots, k-1$ choose $z_j = (e^{2\pi i m \gamma} \omega_j)_{m \in \mathbb{Z}} \in Z$ such that $d(T^j x, z) \leq 2\varepsilon(Z, n)$, and define also $z_k = z_0$ and $\omega_k = \omega_0$. For $0 \leq j < k$ we have

$$\left| e^{2\pi i \gamma} \omega_j - \omega_{j+1} \right| \leq d(T z_j, z_{j+i}) \leq d(T z_j, T^{j+1} x) + d(T^{j+1} x, z_{j+1}) \leq 6\varepsilon(Z, n),$$

and it follows that

$$\left| e^{2\pi i k \alpha} \omega_0 - \omega_0 \right| \leq \sum_{j=0}^{k-1} \left| e^{2\pi i j \gamma} \omega_j - e^{2\pi i (j+1)\gamma} \omega_{j+1} \right| \leq 6 k \varepsilon(Z, n).$$

However, it is well-known [25] that there exists $\delta > 0$ such that $|e^{2\pi i m \alpha} - 1| \geq \delta/m$ for every $m \in \mathbb{N}$, and we deduce that $\varepsilon(Z, n) \geq \delta/6k^2 \geq \delta/6n^2$.

We conclude that if $\mathsf{A} \subset \operatorname{Mat}_d(\mathbb{C})$ is some compact set of matrices which is isometric to S^1, then there exists a minimal invariant set $Z \subset \mathsf{A}^{\mathbb{Z}}$ such that $\varepsilon(Z, n)$ is not $o(n^{-2})$. In particular, the method of Theorem 1.2 is in this case not strong enough even to show that

$$\left| \varrho(\mathsf{A}) - \max_{1 \leq k \leq n} \varrho_k^-(\mathsf{A}) \right| = O\left(\frac{1}{n^{2\alpha}} \right),$$

where α is the Hölder exponent of the function P given by Theorem 2.2. Since $\alpha > 0$ is not explicitly known this estimate would anyway be inferior to the estimate of

J. Bochi described in the introduction. If we wish to achieve further progress using the methods of the present article, therefore, the key step must be to show that for a given set $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ there is an extremal norm $\|\cdot\|$ for which the set

$$(14) \qquad Y = \left\{ x \in \mathsf{A}^{\mathbb{Z}} \colon \varrho(\mathsf{A})^{-n} \| \mathcal{A}(x, n) \| = 1 \ \forall \ n \in \mathbb{N} \right\}$$

contains a minimal set Z such that the quantity $\varepsilon(Z, n)$ decreases with some specified rapidity as a function of n.

It should be remarked that the explicit structure of the set Y defined in (14) is for the most part unknown, and so the range of minimal sets Z which may be contained in such a set Y could in principle be quite limited, potentially leading to improved estimates in Theorem 1.2. Indeed, the the *finiteness conjecture* of J. Lagarias and Y. Wang, proposed in [33], was equivalent to the statement that Y must always contain a periodic orbit. The existence of counterexamples to the finiteness conjecture was established by T. Bousch and J. Mairesse [9], with a simpler argument subsequently being given in [4]. At present, the only well-understood examples of sets A in which Y does not contain a periodic orbit have the property that the orbits in Y are "Sturmian" or "balanced" [9]. When Z consists of Sturmian orbits one may show that $\varepsilon(Z, n)$ decreases exponentially as a function of n, and in particular the arguments used in this article could be applied to obtain an exponential estimate in Theorem 1.2 in this special case.

8. Appendix: Proof of the semi-uniform subadditive ergodic theorem

The proof given below is a condensed exposition of [45], though the hypotheses are slightly weaker and the conclusion slightly stronger. Lemma 8.1 below is a mildly strengthened version of [45, Theorem 1.9]; that result generalises a lemma of M. Herman [26, p.487], which in turn generalises a well-known theorem of Oxtoby [38].

Lemma 8.1. *Let $T \colon X \to X$ be a continuous map of a compact metric space, and let $f \colon X \to \mathbb{R} \cup \{-\infty\}$ be upper semi-continuous. Then*

$$\lim_{n \to \infty} \sup_{x \in X} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \sup_{\mu \in \mathcal{M}_T} \int f \, d\mu.$$

Proof. It is easy to show that the former quantity is an upper bound for the latter. To show the reverse direction, suppose that $(x_n)_{n=1}^{\infty}$ satisfies $(1/n) \sum_{k=0}^{n-1} f(T^k x_n) \geq \lambda$ for infinitely many $n \in \mathbb{N}$. Then using Lemma 3.1 and the compactness of $\mathcal{M}$ we may choose a weak-* limit point μ of the sequence of measures (μ_n) given by $\mu_n = (1/n) \sum_{k=0}^{n-1} \delta_{T^k x_n}$ having the property that $\int f \, d\mu \geq \lambda$. Since clearly $|\int g \, d\mu_n - \int (g \circ T) \, d\mu_n| \to 0$ for every continuous g we have $\mu \in \mathcal{M}_T$. $\qquad \square$

Lemma 8.2. *Let Z be a compact topological space, and let $(g_n)_{n=1}^{\infty}$ be a sequence of upper semi-continuous functions from Z into $\mathbb{R} \cup \{-\infty\}$ such that $(g_n(x))_{n=1}^{\infty}$ is subadditive for every $x \in Z$. Then*

$$(15) \qquad \lim_{n \to \infty} \sup_{z \in Z} \frac{1}{n} g_n(z) = \sup_{z \in Z} \lim_{n \to \infty} \frac{1}{n} g_n(z).$$

Proof. Let $\lambda > \sup_{z \in Z} \lim_{n \to \infty} (1/n) g_n(z)$. For each $z \in Z$ there exists $n_z > 0$ such that $(1/n_z) g_{n_z}(z) < \lambda$, and by upper semi-continuity there is an open neighbourhood U_z of z such that $(1/n_z) g_{n_z}(y) < \lambda$ for all $y \in U_z$. Clearly $\{U_z \colon z \in$

$Z\}$ is an open cover of Z and so we may passing to a finite subcover to deduce that there exist open sets $U_1, \ldots, U_d$ covering Z and integers $n_1, \ldots, n_d$ such that if $z \in U_i$ then $(1/n_i)g_{n_i}(z) < \lambda$. Now take $\hat{n} = \prod_{i=1}^{d} n_i$ and for convenience define $m_i = \hat{n}/n_i \in \mathbb{N}$. If $z \in Z$, then choosing i such that $z \in U_i$ we obtain

$$\frac{1}{\hat{n}}g_{\hat{n}}(z) \leq \frac{1}{m_i} \sum_{k=0}^{m_i-1} \frac{1}{n_i} f_{n_i}(z) < \lambda$$

whence $\sup_{z \in Z}(1/\hat{n})g_{\hat{n}}(z) < \lambda$. Using subadditivity we deduce

$$\lim_{n\to\infty} \sup_{z \in Z} \frac{1}{n}g_n(z) = \inf_{n\geq 1} \sup_{z \in Z} \frac{1}{n}g_n(z) < \lambda,$$

and taking the infimum over λ gives one direction of inequality in (15). The reverse inequality is straightforward: for any $y \in Z$ it is clear that

$$\lim_{n\to\infty} \frac{1}{n}g_n(y) \leq \lim_{n\to\infty} \sup_{z \in Z} \frac{1}{n}g_n(z),$$

and taking the supremum over y yields the required result. $\qquad\square$

Proof of Theorem 3.4. Choose any real number $\lambda > \sup_\mu \lim_n (1/n) \int f_n \, d\mu$. For each $n \in \mathbb{N}$ define a function $g_n \colon \mathcal{M}_T \to \mathbb{R} \cup \{-\infty\}$ by $g_n(\mu) = \int f_n \, d\mu$. By Lemma 3.1 this function is upper semi-continuous, and clearly $(g_n(\mu))_{n=1}^\infty$ is subadditive for every μ. By Lemma 8.2 we obtain

(16) $$\lim_{n\to\infty} \sup_{\mu \in \mathcal{M}_T} \frac{1}{n} \int f_n \, d\mu = \sup_{\mu \in \mathcal{M}_T} \lim_{n\to\infty} \frac{1}{n} \int f_n \, d\mu$$

and it follows that there exists $n_1 > 0$ such that $(1/n_1) \int f_{n_1} \, d\mu < \lambda$ for all $\mu \in \mathcal{M}_T$. Let $M = \sup f_1$. Applying Lemma 8.1 to f_{n_1} it follows that for all sufficiently large integers n_2 we have uniformly for each $x \in X$

$$n_1 f_{n_1 n_2}(x) \leq \sum_{r=0}^{n_1-1} \left(f_r(x) + f_{n_1-r}\left(T^{n_1(n_2-1)+r}x\right) + \sum_{q=0}^{n_2-2} f_{n_1}\left(T^{qn_1+r}x\right) \right)$$

$$\leq M n_1^2 + \sum_{k=0}^{n_1(n_2-1)-1} f_{n_1}\left(T^k x\right) < M n_1^2 + n_1^2(n_2-1)\lambda$$

where we have used the notation $f_0 \equiv 0$ to simplify the presentation. Hence,

$$\inf_{r\geq 1} \sup_{x \in X} \frac{1}{r} f_r(x) \leq \lim_{n_2\to\infty} \frac{M n_1^2 + n_1^2(n_2-1)\lambda}{n_1^2 n_2} = \lambda.$$

Taking the infimum over λ and using subadditivity we obtain

(17) $$\lim_{n\to\infty} \sup_{x \in X} \frac{1}{n}f_n(x) = \inf_{n\geq 1} \sup_{x \in X} \frac{1}{n}f_n(x) \leq \sup_{\mu \in \mathcal{M}_T} \lim_{n\to\infty} \frac{1}{n} \int f_n \, d\mu.$$

By Lemma 3.5 there exists an ergodic measure ν which attains this last supremum. Applying the subadditive ergodic theorem it follows that

(18) $$\sup_{\mu \in \mathcal{M}_T} \lim_{n\to\infty} \frac{1}{n} \int f_n \, d\mu \leq \sup_{x \in X} \limsup_{n\to\infty} \frac{1}{n}f_n(x).$$

Since for every $z \in X$ we clearly have

$$\limsup_{n\to\infty} \frac{1}{n}f_n(z) \leq \lim_{n\to\infty} \sup_{x \in X} \frac{1}{n}f_n(x)$$

we deduce

$$
(19) \qquad \sup_{x \in X} \limsup_{n \to \infty} \frac{1}{n} f_n(x) \le \lim_{n \to \infty} \sup_{x \in X} \frac{1}{n} f_n(x).
$$

Combining (16), (17), (18) and (19) serves to complete the proof.

9. Acknowledgments

References

1. N. E. Barabanov, *On the Lyapunov exponent of discrete inclusions. I*, Avtomat. i Telemekh. (1988), no. 2, 40–46. MR MR940263 (89e:34025)
2. Marc A. Berger and Yang Wang, *Bounded semigroups of matrices*, Linear Algebra Appl. **166** (1992), 21–27. MR MR1152485 (92m:15012)
3. Vincent D. Blondel and Yurii Nesterov, *Computationally efficient approximations of the joint spectral radius*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 1, 256–272 (electronic). MR MR2176820 (2006k:15027)
4. Vincent D. Blondel, Jacques Theys, and Alexander A. Vladimirov, *An elementary counterexample to the finiteness conjecture*, SIAM J. Matrix Anal. Appl. **24** (2003), no. 4, 963–970 (electronic). MR MR2003315 (2004g:15010)
5. Jairo Bochi, *Inequalities for numerical invariants of sets of matrices*, Linear Algebra Appl. **368** (2003), 71–81. MR MR1983195 (2004f:15034)
6. Nicolas Bourbaki, *General topology. Chapters 5–10*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 1998, Translated from the French, Reprint of the 1989 English translation. MR MR1726872 (2000h:54001b)
7. Thierry Bousch, *Le poisson n'a pas d'arêtes*, Ann. Inst. H. Poincaré Probab. Statist. **36** (2000), no. 4, 489–508. MR MR1785392 (2001i:37005)
8. Thierry Bousch and Oliver Jenkinson, *Cohomology classes of dynamically non-negative C^k functions*, Invent. Math. **148** (2002), no. 1, 207–217. MR MR1892849 (2003c:37048)
9. Thierry Bousch and Jean Mairesse, *Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture*, J. Amer. Math. Soc. **15** (2002), no. 1, 77–111 (electronic). MR MR1862798 (2002j:49008)
10. Julien Brémont, *Entropy and maximizing measures of generic continuous functions*, C. R. Math. Acad. Sci. Paris **346** (2008), no. 3-4, 199–201. MR MR2393641
11. Xavier Bressaud and Anthony Quas, *Rate of approximation of minimizing measures*, Nonlinearity **20** (2007), no. 4, 845–853. MR MR2307883 (2008f:37003)
12. D. Collier and I. D. Morris, *Approximating the maximum ergodic average via periodic orbits*, Ergodic Theory Dynam. Systems **28** (2008), no. 4, 1081–1090. MR MR2437220
13. G. Contreras, A. O. Lopes, and Ph. Thieullen, *Lyapunov minimizing measures for expanding maps of the circle*, Ergodic Theory Dynam. Systems **21** (2001), no. 5, 1379–1409. MR MR1855838 (2002i:37038)
14. J.-P. Conze and Y. Guivarc'h, *Croissance des sommes ergodiques*, unpublished manuscript, circa 1990.
15. Ingrid Daubechies and Jeffrey C. Lagarias, *Sets of matrices all infinite products of which converge*, Linear Algebra Appl. **161** (1992), 227–263. MR MR1142737 (93f:15006)
16. _____, *Two-scale difference equations. II. Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal. **23** (1992), no. 4, 1031–1079. MR MR1166574 (93g:39001)
17. Jean Marie Dumont, Nikita Sidorov, and Alain Thomas, *Number of representations related to a linear recurrent basis*, Acta Arith. **88** (1999), no. 4, 371–396. MR MR1690457 (2000c:11030)
18. L. Elsner, *The generalized spectral-radius theorem: an analytic-geometric proof*, Proceedings of the Workshop "Nonnegative Matrices, Applications and Generalizations" and the Eighth Haifa Matrix Theory Conference (Haifa, 1993), vol. 220, 1995, pp. 151–159. MR MR1334574 (96e:15010)

19. Gary Froyland, Simon Lloyd, and Anthony Quas, *Coherent structures and isolated spectrum for perron-frobenius cocycles*, preprint, 2008.

20. I. C. Gohberg and M. G. Kreĭn, *Introduction to the theory of linear nonselfadjoint operators*, Translated from the Russian by A. Feinstein. Translations of Mathematical Monographs, Vol. 18, American Mathematical Society, Providence, R.I., 1969. MR MR0246142 (39 #7447)

21. Gustaf Gripenberg, *Computing the joint spectral radius*, Linear Algebra Appl. **234** (1996), 43–60. MR MR1368770 (97c:15043)

22. N. Guglielmi, F. Wirth, and M. Zennaro, *Complex polytope extremality results for families of matrices*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 3, 721–743 (electronic). MR MR2208331 (2007b:93106)

23. N. Guglielmi and M. Zennaro, *On the zero-stability of variable stepsize multistep methods: the spectral radius approach*, Numer. Math. **88** (2001), no. 3, 445–458. MR MR1835466 (2002d:65066)

24. Leonid Gurvits, *Stability of discrete linear inclusion*, Linear Algebra Appl. **231** (1995), 47–85. MR MR1361100 (96i:93056)

25. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, sixth ed., Oxford University Press, Oxford, 2008, Revised by D. R. Heath-Brown and J. H. Silverman. MR MR2445243

26. Michael-R. Herman, *Une méthode pour minorer les exposants de Lyapounov et quelques exemples montrant le caractère local d'un théorème d'Arnol′d et de Moser sur le tore de dimension* 2, Comment. Math. Helv. **58** (1983), no. 3, 453–502. MR MR727713 (85g:58057)

27. Karl Heinrich Hofmann and Paul S. Mostert, *Elements of compact semigroups*, Charles E. Merr ll Books, Inc., Columbus, Ohio, 1966. MR MR0209387 (35 #285)

28. Oliver Jenkinson, *Ergodic optimization*, Discrete Contin. Dyn. Syst. **15** (2006), no. 1, 197–224. MR MR2191393

29. Boris Kalinin, *Livsic theorem for matrix cocycles*, preprint, 2008.

30. Tosio Kato, *Perturbation theory for linear operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995, Reprint of the 1980 edition. MR MR1335452 (96a:47025)

31. V. S. Kozyakin, *Algebraic unsolvability of a problem on the absolute stability of desynchronized systems*, Avtomat. i Telemekh. (1990), no. 6, 41–47. MR MR1071607 (91e:93065)

32. Ulrich Krengel, *Ergodic theorems*, de Gruyter Studies in Mathematics, vol. 6, Walter de Gruyter & Co., Berlin, 1985, With a supplement by Antoine Brunel. MR MR797411 (87i:28001)

33. Jeffrey C. Lagarias and Yang Wang, *The finiteness conjecture for the generalized spectral radius of a set of matrices*, Linear Algebra Appl. **214** (1995), 17–42. MR MR1311628 (95k:15038)

34. Mohsen Maesumi, *An efficient lower bound for the generalized spectral radius of a set of matrices*, Linear Algebra Appl. **240** (1996), 1–7. MR MR1387282 (97b:15027)

35. ———, *Calculating joint spectral radius of matrices and Hölder exponent of wavelets*, Approximation theory IX, Vol. 2 (Nashville, TN, 1998), Innov. Appl. Math., Vanderbilt Univ. Press, Nashville, TN, 1998, pp. 205–212. MR MR1744409

36. Bruce E. Moision, Alon Orlitsky, and Paul H. Siegel, *On codes that avoid specified differences*, IEEE Trans. Inform. Theory **47** (2001), no. 1, 433–442. MR MR1820392 (2001k:94084)

37. I. D. Morris, *A sufficient condition for the subordination principle in ergodic optimization*, Bull. Lond. Math. Soc. **39** (2007), no. 2, 214–220. MR MR2323451 (2008c:37052)

38. John C. Oxtoby, *Ergodic sets*, Bull. Amer. Math. Soc. **58** (1952), 116–136. MR MR0047262 (13,850e)

39. Pablo A. Parrilo and Ali Jadbabaie, *Approximation of the joint spectral radius of a set of matrices using sum of squares*, Hybrid systems: computation and control, Lecture Notes in Comput. Sci., vol. 4416, Springer, Berlin, 2007, pp. 444–458. MR MR2363632 (2008m:15027)

40. M. S. Raghunathan, *A proof of Oseledec's multiplicative ergodic theorem*, Israel J. Math. **32** (1979), no. 4, 356–362. MR MR571089 (81f:60016)

41. Gian-Carlo Rota and Gilbert Strang, *A note on the joint spectral radius*, Nederl. Akad. Wetensch. Proc. Ser. A 63 = Indag. Math. **22** (1960), 379–381. MR MR0147922 (26 #5434)

42. David Ruelle, *Ergodic theory of differentiable dynamical systems*, Inst. Hautes Études Sci. Publ. Math. (1979), no. 50, 27–58. MR MR556581 (81f:58031)

43. Sebastian J. Schreiber, *On growth rates of subadditive functions for semiflows*, J. Differential Equations **148** (1998), no. 2, 334–350. MR MR1643183 (2000a:37004)

44. Mau-Hsiang Shih, Jinn-Wen Wu, and Chin-Tzong Pang, *Asymptotic stability and generalized Gelfand spectral radius formula*, Linear Algebra Appl. **252** (1997), 61–70. MR MR1428628 (97k:15028)

45. R. Sturman and J. Stark, *Semi-uniform ergodic theorems and applications to forced systems*, Nonlinearity **13** (2000), no. 1, 113–143. MR MR1734626 (2000m:37041)

46. John N. Tsitsiklis and Vincent D. Blondel, *The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate*, Math. Control Signals Systems **10** (1997), no. 1, 31–40. MR MR1462278 (99h:65238a)

47. Peter Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York, 1982. MR MR648108 (84e:28017)

48. Fabian Wirth, *The generalized spectral radius and extremal norms*, Linear Algebra Appl. **342** (2002), 17–40. MR MR1873424 (2003g:15025)

49. Guocheng Yuan and Brian R. Hunt, *Optimal orbits of hyperbolic systems*, Nonlinearity **12** (1999), no. 4, 1207–1224. MR MR1709845 (2000g:37029)